

# PATENT ABSTRACTS OF JAPAN

(1) Publication number: 10-293860

(43) Date of publication of application: 04.11.1998

(5) Int.Cl. 601 13/00  
610 3/00

(21) Application number: 09-216377 (71) Applicant: NIPPON TELEGR & TELEPH CORP <NTT>

(22) Date of filing: 11.08.1997 (72) Inventor: MATSUURA MICHIAKI  
KURA TSUNEKO  
OSHIMA TAKASHI  
WATANABE NOBUYUKI  
KANAYAMA HIDEAKI

(30) Priority  
Priority number: 09 38760 Priority date: 24.02.1997 Priority country: JP

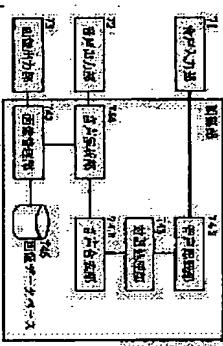
(54) PERSON IMAGE DISPLAY METHOD AND DEVICE USING VOICE DRIVE

(57) Abstract:  
PROBLEM TO BE SOLVED: To artificially attain an almost natural conversation state with no use of any special device, etc., by changing at least one of

shakes of a mouth, head, etc., of a person face image based on the loudness of human voices which are obtained in sequence and at each fixed time interval.

SOLUTION: A voice analysis part 744 picks up the changes of voices received from a voice synthesizing part 743 at each prescribed time interval and calculates the mean value of loudness of voices, etc. Then the time series information on each of mean value of voice loudness divided at each time interval of voices is outputted to an image management part 745. The part 745 successively selects the prescribed proper images out of an image data base in response to each information and outputs these images to an image output part 73 to successively display them.

These images primarily show an answering person to the voices of a speaker, and the mouths, attitudes, etc., of one or more persons are operated in accordance with the voices produced from a voice output part 72.



(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

特開平10-293860

(43) 公開日 平成10年(1998)11月4日

(51) Int. Cl. G 06 T 13/00 G 06 F 15/62 G 10 L 3/00

魔別記号 F 1 G 06 F 15/62 3 4 0 D

G 10 L 3/00 S

(21) 出願番号 特願平9-216377  
(22) 出願日 平成9年(1997)9月11日  
(31) 優先権主張番号 特願平9-38760  
(32) 優先日 平9(1997)2月24日  
(33) 優先権主張国 日本 (JP)

(71) 出願人 日本電信電話株式会社  
(72) 発明者 松浦 進明  
東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(72) 発明者 大島 孝  
東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内  
(74) 代理人 弁理士 秋田 收容  
最終回に続く

(54) 【発明の名稱】音声駆動を用いた人物画像表示方法およびその装置

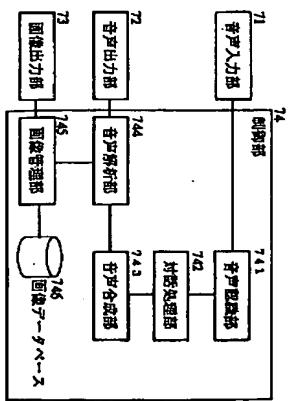
(57) 【要約】

【課題】専門の技術あるいは装置を必要としなくともほぼ自然的に対話をしている状態を実現的に実現でき

る。

【解決手段】コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示方法において、前記人物の発する音声の一一定時間隔ごとに最大尋ねられる各音の大きさに基づいて、該人物の顔画像における口、顎の縮あるいはうなずき、目のまばたきの少なくとも一つを変化させる。

図 1



(11) 特許出願公開番号

【特許請求の範囲】  
【請求項1】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声運動を用いた人物画像表示方法において、

前記人物の発する音声の一一定時間隔ごとに順次得られる各音声の大きさに基づいて、該人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とする音声運動を用いた人物画像表示方法。

【請求項2】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声運動を用いた人物画像表示装置において、

前記人物の発する音声の一一定時間隔ごとに順次得られる各音声の大きさに基づいて、該人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とする音声運動を用いた人物画像表示方法。

【請求項3】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声運動を用いた人物画像表示装置において、

前記人物の発する音声の一一定時間隔ごとに順次得られる各音声の大きさに基づいて、該人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とする音声運動を用いた人物画像表示方法。

【請求項4】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声運動を用いた人物画像表示装置において、

前記人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つが変化された各画像

を前記人物の発する音声の一一定時間隔ごとに順次得られることを特徴とする音声運動を用いた人物画像表示方法。

【請求項5】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によ

て、前記人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つが変化された各画像

を前記人物の発する音声の一一定時間隔ごとに順次得られることを特徴とする音声運動を用いた人物画像表示方法。

【請求項6】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によ

て、前記人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つが変化された各画像

を前記人物の発する音声の一一定時間隔ごとに順次得られることを特徴とする音声運動を用いた人物画像表示方法において、

前記人物の発する音声から一定時間隔毎に得られる有音/無音の判定に応じて、口の大きさ、目の瞬き、頭の揺れ、うなずきの少なくとも一つを制御することを特徴とする音声運動を用いた人物画像表示方法。

【請求項7】 コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によ

て、前記人物の発する音声から一定時間隔毎に得られる有音/無音の判定に応じて、口の大きさ、目の瞬き、頭の揺れ、うなずきの少なくとも一つを制御することを特徴とする音声運動を用いた人物画像表示装置において、

前記人物の発する音声から一定時間隔毎に得られる有音/無音の判定に応じて、口の大きさ、目の瞬き、頭の揺れ、うなずきの少なくとも一つを制御することを特徴とする音声運動を用いた人物画像表示方法。

【請求項8】 有音判定の際に、予め定められた周期あるいはランダムな時間隔で口の大きさを変化させ、無音判定の際に、予め定められた周期あるいはランダムな時間隔で目を閉じることを特徴とする音声運動を用いた人物画像表示方法。

【請求項9】 有音判定の際に、予め定められた周期あるいはランダムな時間隔で頭を動かさせることを特徴とする音声運動を用いた人物画像表示方法。

【請求項10】 有音判定の際に、予め定められた周期あるいはランダムな時間隔で頭を動かさせることを特徴とする音声運動を用いた人物画像表示方法。

【請求項11】 無音判定の際に、予め定められた周期あるいはランダムな時間隔で顔方向に頭を動かせるか、首を傾げか、首を机に置くことを行なう装置を用いた音声運動を用いた人物画像表示方法。

【請求項12】 人物の顔の大きさに基づいて、該人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とする装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は音声運動を用いた人物画像表示方法およびその装置に係り、たとえばコンピュータヒンタラクタイプに対話を行うシステムにおいて、音声情報を用いて人物の画像の顔部分を変形させることにより、対話をしている状態を観察的に実現する方法および装置に関する。

【0002】

【従来の技術】 この種の方法および装置としては、たとえばニューラルネットワークを用いて口の形を決定する方法が知られている(文献:官下、佐藤、坂口、森島「仮想人物との対話をするための音声から画像への実時間メディア変換システムの研究」信学技法誌95-62,1996-03)。

【0003】 すなわち、音声の特徴パラメータである発音者の声声特性と映像特性を表しているLPCケープストラムの係数から口形情報をニューラルネットワークを用いた人物画像表示装置。

【請求項1】 コンピュータに表示されている人物の発

用いてを行い、これにより実際の口の形に近い表示を可能としている。

【0004】

【発明が解決しようとする課題】 しかしながら、このようないニューラルネットワークを用いた口形の決定に要する処理は、高度の専門の技術を要し、さらに、高度の専用の装置を必要とするという問題が指摘されるに到つて

いる。

【0005】 そのため、このような問題を解決した結果、たとえばアニメーションや吹き替えの映画等を鑑賞する場合に明らかなように、必ずしも口形と実際の音声とが一致しない場合、話し始め口を開き話し終わ

りで口が閉じることにより、ほとんどの観客が不自然さを感じない、実感を考慮すると、必ずしも専門の技術あるいは装置を必要としなくともほんの自然に対話をしている状態を観察的に実現できることが判明した。

【0006】 本発明は、このようないニューラルネットを用いた装置を提供することにある。

【0007】

【課題を解決するための手段】 本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、下記の通りである。

【0008】 すなわち、コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声運動を用いた人物画像表示方法。

【請求項1】 無音判定の際に、予め定められた周期あるいはランダムな時間隔で顔方向に頭を動かせるか、首を傾げか、首を机に置くことを行なう装置を用いた音声運動を用いた人物画像表示方法。

【0009】 すなわち、音声運動を用いた人物画像表示方法によれば、人物の表情において、たとえば

顔の場合は、口の大きさが、さらに顎の動きで口が閉じるようになり、また声の大きさに分類したことにより、対話をしている状態を観察的に実現する

ことができる。この結果、その声の大きさを変化させることによって、その声の大きさが不自然さを感じることなく隠れることができるようにになる。

【0010】 したがって、特に、複数種手を用いる必要がないことから、専門の技術あるいは装置を必要とすることなくほんの自然に対話をしている状態を観察的に実現するようになる。

【0011】

【発明の詳細な説明】 以下、本発明による音声運動を用いた人物画像表示方法および装置の実施例について図面を用いて説明する。

【0012】 実施例1. 図1は、本発明による音声運動を用いた人物画像表示装置の一実施例を示すプロック図

である。

【0013】 同図において、コンピュータヒンタラクタイプに對話を行う対話者の音声を入力する音声入力部7-1があり、この音声入力部7-1からの信号はコンピュータ側の制御部7-4に取り込まれるようになっている。

【0014】 制御部7-4では、まず、音配音声入力部7-1からの信号を音声認識部7-4-1によって、その音声を認識し、かつその音楽の意味を判断するようになってい

る。

【0015】 そして、対話処理部7-4-2では対話者の意

図を読み取り、かつ、それに対する返答を判断するよう

になっている。

【0016】 音声合成部7-4-3では、前記対話処理部7-4-2によって判断された返答に相当する音声を作成する

ようになっている。この音声は強さの変化のない単調なものではなく、あたかも実際の感情を有する人間が話すようにその音楽の意味に応じて強弱(アクセントおよび

イントネーション)が付けられたものとなっている。

【0017】 そして、この音声合成部7-4-3から出力され、この音声出力部7-2から前記返答に相当する音声が

発せられるようになっている。

【0018】 前記音声解析部7-4-4では、音声合成部7-4-3からの音声に對してたとえばT-1の時間隔ごとに

音声の変化を取りだし、その強さの平均値の算出、あ

るいはその他の必要な演算処理を行なっている。

【0019】 そして、この音声解析部7-4-4で得られた音声の時間毎に区切られた各平均値等の時間的各情

况は音声管理部7-4-5に由来され、この音声管理部7-4-5では、前記それぞれの情報に応じて画像データベース7-4-6から予め定められた適当な画像が順次選択され、この画像は画像出力部7-3に出力されて順次表示されるようになっている。

【0020】 この画像は、前記対話者の音声に対して返答をする人物の画像(一人の場合は、複数の場合もあるし、複数の場合もある)が主となり、その人物の口あるいはそばり等は、前記音声出力部7-2から差しられる音声に追従して(同じタイミングで)動くようになっている。

【0021】 次に、前記音声解析部7-4-4、音声管理部7-4-5、および画像データベース7-4-6の各機能を、基

本アルゴリズムおよびそれをさらに拡張された拡張アルゴリズムによって以下説明する。

【0022】 基本アルゴリズムとして、まず、説明の段

のどなっている。

【0023】 そして、音声合成部7-4-3からの音声信号を音声解析部7-4-4によって、図2に示すように、ある

一定時間の間隔T-1ごとに音声の大きさの平均値S-0を





[10106] また、このような構成は、たとえばインターネットフォンのように、パソコンで通信しながらお互いの音声を用いて会話するようにすることもできる。この場合、音声だけでなく、画像データベース 7.4 に登録してある顔画像（たとえば話をする相手の顔画像）を自分のパソコンの画面上に表示する。データ受信部 7.4 から送られてきた音声の強弱などの情報を音声解釈部 7.4 で求め、その結果を受けて画像データベース 7.4 から適当な画像が順次選択され、画像出力部 7.3 に出力される。

[10107] 以上、上述した各実施例によれば、人物の表情において、たとえばその口の場合は、話す始めで口が開き話し、さらに終わりで口が閉じるようになり、また声の大きさ等で分類した数に応じてそれらの間で口の開きの大きさを変化させることができることから、ほとんどの対話者が不自然さを感じることなく鍛錬することができるようになる。

[10108] したがって、特に、複雑な手法を用いる必要がないことから、専門の技術あるいは装置を必要とするところなくほぼ自然的に対話をしている状態を実験的に実現できるようになる。

[10109] なお、上述した各実施例では、いわゆる本アルゴリズムおよび拡張アルゴリズムにしたがって動作するように改良したものの、これに限定されることはなく、それらの一方にしたがって動作できるようにしておきたいことはいうまでもない。

【発明の効果】 以上説明したことから明らかなるように、本発明による音声認識を用いた人物画像表示方法および装置によれば、専門の技術あるいは装置を必要としなくてもほぼ自然的に対話をしている状態を擬似的に実現できるようになる。

【図1】本発明による音声駆動を用いた人物画像表示装置の一実施例を示すブロック図である。

【図2】本発明の基本アルゴリズムの一実施例を示す説明図である。

【図3】本発明の基本アルゴリズムの一実施例を示す説明図である。

【図4】本発明の基本アルゴリズムの一実施例を示す説明図である。

【図5】本発明の並進アルゴリズムの一実施例を示す説明図である。

【図6】本発明の並進アルゴリズムの一実施例を示す説明図である。

【図7】本発明の並進アルゴリズムの一実施例を示す説明図である。

【図8】本発明の一実施例におけるアルゴリズムを実験的に試みた際の図像を示す図である。

【図9】本発明の並進アルゴリズムのさらに並進された実施例を示すタイムチャートで、図10とで一つの図を構成している。

【図10】本発明の並進アルゴリズムのさらに並進された実施例を示すタイムチャートで、図9とで一つの図を構成している。

【図11】図9および図10に示すタイムチャートを一つの図で、図12とで一つの図を構成している。

【図12】図9および図10に示すタイムチャートを一つの図で、図11とで一つの図を構成している。

【図13】図11および図12に示すフロー図の動作に先立つて予め装置に格納しておくデータを示す説明図である。

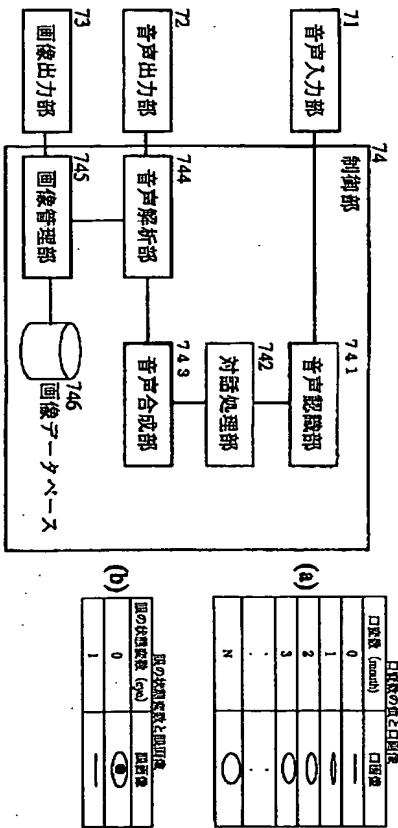
【図14】本発明による音声認識を用いた人物画像表示装置の他の実施例を示すプロック図である。  
【符号の説明】  
30 1.……音声入力部、7.2.……音声出力部、7.3.……画  
像出力部、7.4.……動画部、7.4.1.……音声認識部、7.

…音声解析部、745…画像管理部、746…画像データベース。

三

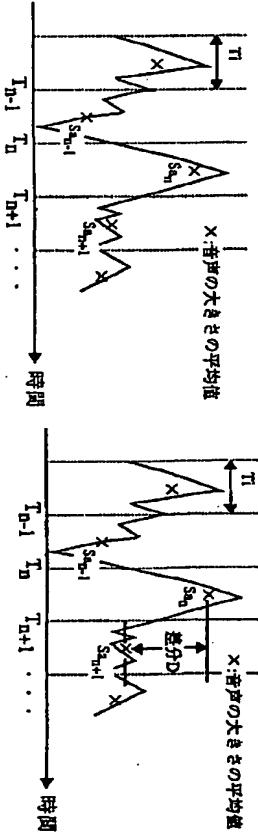
1

11



四

5



619

四

四

四

[図4]

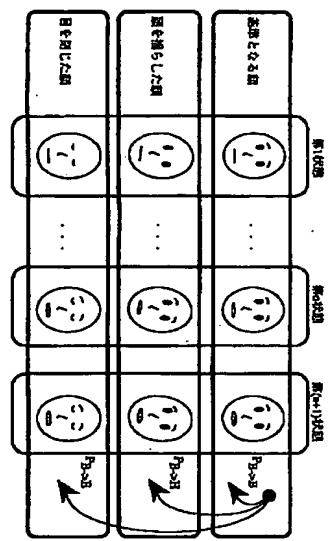
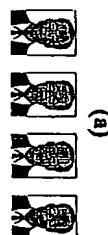
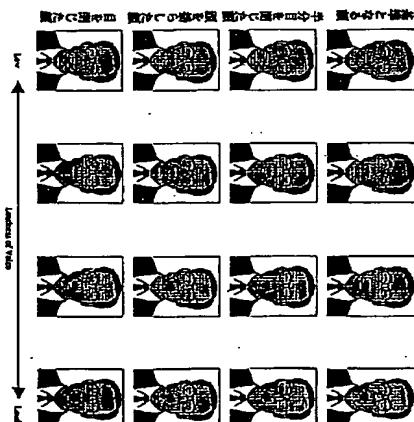


図4

[図8]



(a)

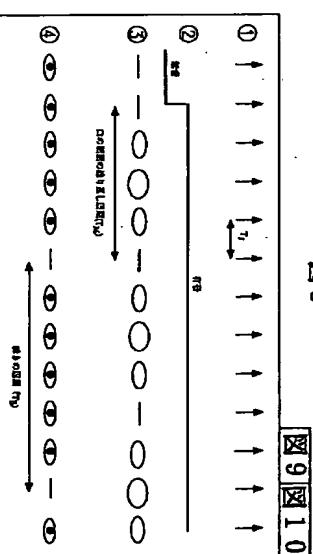


(b)

[図9]

[図9]

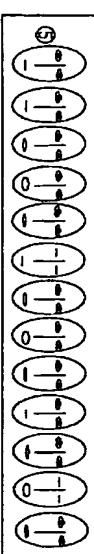
[図10]



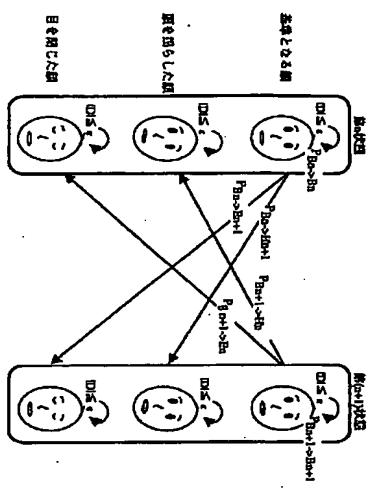
[図9]

[図9]

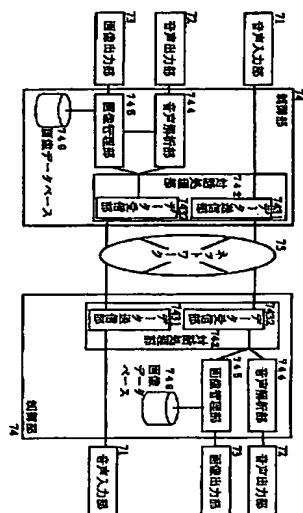
[図10]



[図7]

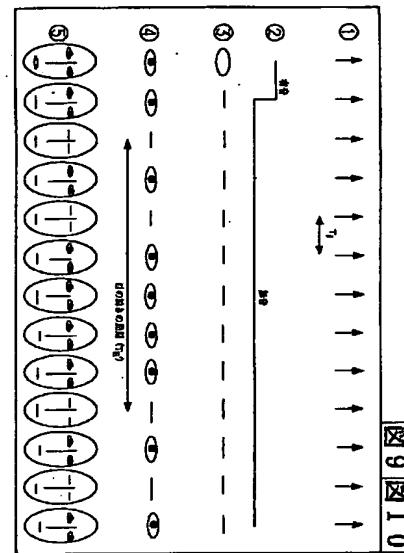


[図7]



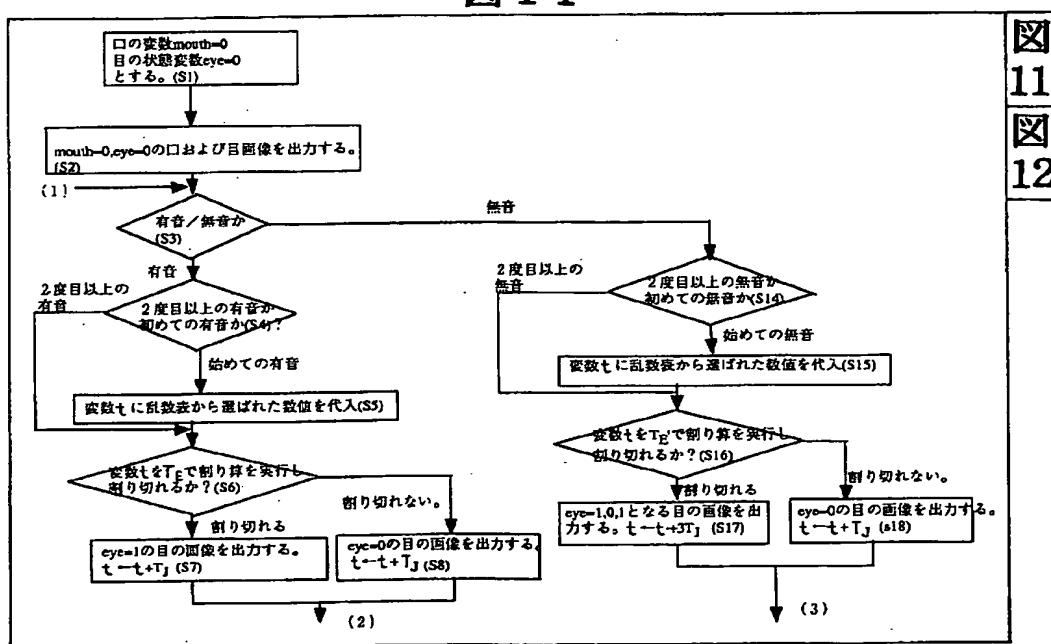
四  
14

141



10

୧୦୫

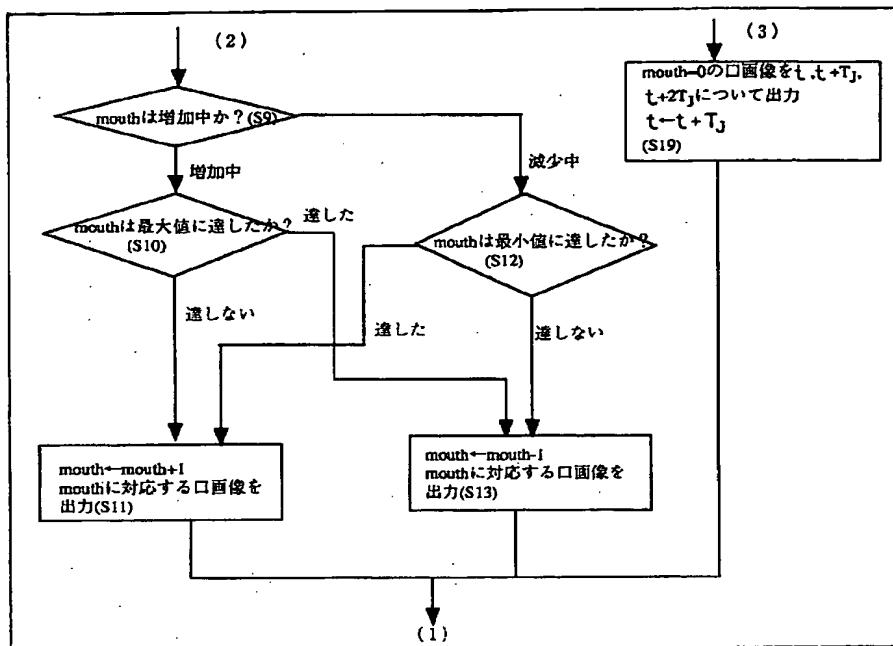


[11]

[図12]



図12



フロントページの動き

(72)発明者 鶴見 健之

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72)発明者 金山 英明

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内